Thomson's formulation of the second law: an exact theorem and limits of its validity

Armen E. Allahverdyan[*], Roger Balian[†] and Theo M. Nieuwenhuizen[**]

[*]*Yerevan Physics Institute, Alikhanian Brothers St. 2, Yerevan 375036, Armenia*
[†]*Service de Physique Théorique, CEA Saclay, 91191 Gif-sur-Yvette cedex, France*
[**]*Institute for Theoretical Physics, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands*

Abstract. Thomson's formulation of the second law - no work can be extracted from a system coupled to a bath through a cyclic process - is believed to be a fundamental principle of nature. For the equilibrium situation a simple proof is presented, valid for macroscopic sources of work. Thomson's formulation gets limited when the source of work is mesoscopic, i.e. when its number of degrees of freedom is large but finite. Here work-extraction from a single equilibrium thermal bath is possible when its temperature is large enough. This result is illustrated by means of exactly solvable models. Finally we consider the Clausius principle: heat goes from high to low temperature. A theorem and some simple consequences for this statement are pointed out.

Introduction

Since one century the laws of thermodynamics no longer appear as basic principles, but as consequences of the laws of matter at the microscopic level. Their derivation from first principles requires an identification of concepts such as temperature, work or heat in terms of the elementary constituents, and an elimination of the irrelevant microscopic variables by means of methods of statistical mechanics. This program has been partially achieved for macroscopic systems in the thermodynamic limit [1, 2]. Here we focus on Thomson's (later Lord Kelvin) formulation of the second law for equilibrium systems. We have two basic goals. The first is to discuss a rigorous version of this formulation and to present a simple proof for it. This version uses a minimal number of assumptions and has a larger region of applicability as compared to the text-book presentation of Thomson's law [1, 2], where it is typically derived from the law of entropy increase. Our second goal is to determine mesoscopic limits of applicability of the formulation. They arise due to the existence of large but not very large number of degrees of freedom (mesocopicity). The special importance of Thomson's formulation comes from the fact that other formulations rely more explicitly upon notions such as entropy or temperature which require rather elaborate identification in terms of microscopic variables, whereas this one concentrates on the concept of work (energy exchange) which can be considered as a primary, microscopic concept, as we show below.

General setup for Thomson's formulation

In its original form Thomson's formulation states that if a thermodynamic system S, which interacts with a single macroscopic thermal bath B at uniform temperature T and with a source of work W, evolves in a closed cycle, it cannot yield work to the latter external system W. In other words, such a process cannot extract energy from the heat bath.

We assume that interaction of B with S is sufficiently weak, yielding a negligible contribution to the energy. (Later in this section we will comment on the situation with strong coupling with the bath). Over sufficiently long times, energy transfer identified with heat may result from this interaction. The bath can therefore set up a canonical equilibrium at the temperature T for S, without interacting directly with the source of work.

In order to separate from each other the effects of the interactions of S with B and with W, we shall consider in the following a cyclic process that takes place in two steps. A canonical thermal equilibrium of S is assumed to have already been established for times $t < 0$. The interaction process that couples S and W takes place between the times $t = 0$ and $t = \tau$. The delay τ is supposed to be sufficiently short so that no heat is exchanged with the bath during this first step of the process. This is possible due to a weak coupling between the system S and the bath B. During the second step, for $t > \tau$, the energy of W is left unchanged, while S returns to its original thermal equilibrium state at temperature T through interaction with the bath B. Altogether, if the cycle is closed as regards the system S, a positive amount of energy should according to Thomson have been transferred from W to S between the times $t = 0$ and $t = \tau$, and this amount should be released from S to B for $t > \tau$. This two-step analysis of the process will allow us to focus first on the coupled evolution of S and W in the time-interval $0 < t < \tau$, and to postpone the forthcoming thermal exchange with the bath.

More precisely, in order to define the *work* done by the source W, we express the Hamiltonian which generates the evolution in this time-interval as

$$\hat{H} = \hat{H}_{\rm S} + \hat{H}_I + \hat{H}_{\rm W}, \tag{1}$$

where $\hat{H}_{\rm S}$ ($\hat{H}_{\rm W}$) depends only on the variables of S (W), with $[\hat{H}_{\rm S}, \hat{H}_{\rm W}] = 0$, and where $\hat{H}_I$ describes the coupling between S and W. To fix ideas we shall first take $\hat{H}_I$ in the form: $\hat{H}_I = \hat{X}\hat{\xi}$. The observable $\hat{X}$ belongs to the system S, and corresponds to a "force" variable such as the pressure for a gas or the magnetic moment for a magnetic material. The conjugate variable $\hat{\xi}$ refers to the system W, and corresponds to a "position" variable such as the volume for a gas or the magnetic field for a magnetic source. Work done by S on W till a moment is identified as the energy yielded by the source W, or equivalently as the decrease of the expectation value $\hat{H}_{\rm W}$ of its Hamiltonian [2]: $W = \langle H_{\rm W} \rangle_0 - \langle H_{\rm W} \rangle_t$.

Thomson's formulation for macroscopic sources

In the thermodynamic limit for the system W (but not necessarily for the system S), the fluctuations in $\hat{\xi}$ are negligible at all relevant times; in relative value they are small

as the inverse square root of the number of microscopic degrees of freedom of W. The operator $\hat{\xi}$ can therefore be replaced by its expectation value $\langle \hat{\xi} \rangle_t = \mathrm{tr}[\hat{\xi} \rho_W(t)] \equiv \xi(t)$, $\rho_W(t) = \mathrm{tr}_S \rho(t)$, where $\rho(t)$ is the total density matrix of S and W, and where $\rho_W(t)$ is the marginal density matrix of W. The evolution of $\xi(t)$ is generated by the Ehrenfest equation: $\hbar \, \mathrm{d}\xi(t)/\mathrm{d}t = \mathrm{i}\langle [\hat{H}, \hat{\xi}] \rangle_t$.

More precisely, if the fluctuation $\langle \Delta \hat{\xi}^2 \rangle_t$, where $\Delta \hat{\xi} = \hat{\xi} - \langle \hat{\xi} \rangle_t$, is negligible *at all times*, the Schwartz inequality $|\langle \Delta \hat{\xi} \, \Delta \hat{Z} \rangle_t|^2 \leq \langle \Delta \hat{\xi}^2 \rangle_t \langle \Delta \hat{Z}^2 \rangle_t$, where $\hat{Z}$ is any operator of S or W, implies that we can make the replacement of the operator $\hat{\xi}$ by its expectation value $\langle \hat{\xi} \rangle_t = \xi(t)$ everywhere, in particular in its Hamiltonian, which thus takes the form $\hat{H} = \bar{H}_S(t) + \hat{H}_W = \hat{H}_S + \hat{X}\,\xi(t) + \hat{H}_W$, where an effective time-dependent Hamiltonian $\bar{H}_S(t)$ appears for the system S. This form of $\hat{H}$ implies the factorization of the evolution operator, and that of the density operator $\rho(t) = \rho_S(t) \otimes \rho_W(t)$ at all times, provided the initial state at $t = 0$ is also factorized. Hence in the Schrödinger picture the density matrix $\rho_S(t)$ of the system S evolves according to the unitary transformation U generated by $\bar{H}_S(t)$.

The equilibrium condition at time $t = 0$ provides the initial density operator of S as

$$\rho_S(0) = \mathrm{e}^{-\bar{H}_S(0)/T}/Z, \qquad Z = \mathrm{tr}\, \mathrm{e}^{-\bar{H}_S(0)/T}. \tag{2}$$

A cyclic transformation is characterized here by the fact that $\xi(t)$ has returned to its initial value at time $t = \tau$: $\xi(\tau) = \xi(0)$, $\quad \bar{H}_S(\tau) = \bar{H}_S(0)$. Owing to the lack of fluctuations in $\hat{\xi}$, this condition is sufficient to ensure that the overall process can be rigorously cyclic at the microscopic level. Indeed, since the state of S is not correlated at the time $t = \tau$ with that of W, it can return later on to its original form (2) at $t = 0$ through interaction with the thermal bath B. The work given by W to S between the times 0 and τ is equal due to: $W = \mathrm{tr}[\rho(\tau)\bar{H}_S(0) - \rho_S(0)\bar{H}_S(0)]$. With help of (2) one gets, using the fact that U is unitary:

$$\frac{W}{T} = \mathrm{tr}[\rho_S(\tau)\ln\rho_S(\tau) - \rho_S(\tau)\ln\rho_S(0)] \geq 0. \tag{3}$$

This expression is nothing but the *relative entropy* of $\rho_S(\tau)$ with respect to $\rho_S(0)$; it is strictly positive if $\rho_S(\tau) \neq \rho_S(0)$ [3].

As shown by (3), the work is positive for a macroscopic source of work W, whether the system S is macroscopic or microscopic. Thus, in the second case this entails an extension of the original Thomson's statement [4, 5], which was verified in specific examples [6, 7]. The irreversibility of the process arises from the lack of symmetry between the times 0 and τ which is introduced by the initial condition (2). Usually, the breaking of time-reversal-invariance is associated with a dissipative process. Here the latter process is the interaction with the heat bath, which will take place after the time τ.

The above treatment can readily be generalized to an interaction with several (not necessarily independent) sources of work, where $\hat{H}_I = \sum_{i=1}^{r} \hat{X}_i \hat{\xi}_i$. Altogether, if the source W of work is a macroscopic system, the lack of fluctuations in the variables $\hat{\xi}_i$, which couple W to the system S, ensures the second law in Thomson's form as soon as S is initially in a canonical equilibrium state.

The fate of Thomson's formulation for mesoscopic sources.

Let us now turn to the situation where W is no longer a macroscopic object. We have to deal with the operator nature of the $\hat{\xi}_i$'s and we can no longer leave aside the Hamiltonian $\hat{H}_W$ of the source of work, when studying the dynamics of S.

Work is still defined as the decrease of the energy of the source. We will again assume that in its initial state the system S is in canonical thermal equilibrium, and is not correlated with W: $\rho(0) = \rho_S(0) \otimes \rho_W(0)$, where $\rho_S(0)$ is given by (2) with $\bar{H}_S(0) = \hat{H}_S + \sum_{i=1}^r \hat{X}_i \langle \hat{\xi}_i \rangle_0$. A detailed discussion on the necessity of this initial conditions is given in [8]. However, the definition of a cyclic process is no longer unambiguous. This condition may indeed be generalized in two different ways, both of which are rather natural [8]. To keep our presentation concise we will here outline only one of the two definitions proposed in [8]. Since the variables $\hat{X}_i$ and $\hat{\xi}_i$ can now be correlated, and since the interaction term $\hat{H}_I$ involves a sum of their products, we search a condition for such a sum of products. We notice that in the considered mesoscopic case the states of S and W do become correlated for $t > 0$. Since now both $\rho(t)$ and $\hat{H}_I$ irreducibly live in the common Hilbert space of S and W, the most standard way to define energy exchange between two systems is to require that the average interaction Hamiltonian does not contribute to the total energy budget, i.e., it is zero both initially and finally. For our situation the proper interaction Hamiltonian is $\sum_{i=1}^r \hat{X}_i(\hat{\xi}_i - \langle \hat{\xi}_i \rangle_0)$, since its average is zero at the initial time $t = 0$. Requiring that this average is also equal to zero at some final time τ, we get for a cyclic process with the duration τ:

$$\sum_{i=1}^r \langle \hat{X}_i \hat{\xi}_i \rangle_\tau = \sum_{i=1}^r \langle \hat{X}_i \rangle_\tau \langle \hat{\xi}_i \rangle_0. \tag{4}$$

For a large system W we recover the macroscopic condition for $r = 1$, since the lack of fluctuations in $\hat{\xi}$ implies the lack of correlations with $\hat{X}$: $\langle \hat{X} \hat{\xi} \rangle_t \simeq \langle \hat{X} \rangle_t \langle \hat{\xi} \rangle_t$.

Work in the short-time limit. We shall discuss for (4) the sign of the work integrated over the duration τ of a closed cycle, in the limit where τ is sufficiently small so that we can expand W in powers of τ. This limit is especially interesting, since mesoscopic contributions to the work can overcome the main thermodynamic term: as we shown [8], the thermodynamic contribution to the work, evaluated by neglecting the fluctuations of $\hat{\xi}_i$'s and given by (3), scales as τ^6 for small τ, whereas in the mesoscopic case $W \sim \tau^2$. Expanding the Heisenberg equations of motion one gets for the work [8]:

$$\frac{2W_2}{\tau^2} = \sum_{i,k=1}^r \langle \{\Delta\hat{\xi}_i, \Delta\hat{\xi}_k\}_+ \rangle \sum_{\alpha<\beta} (p_\alpha - p_\beta)(h_\beta - h_\alpha)\Re\{\langle\alpha|\hat{X}_i|\beta\rangle\langle\beta|\hat{X}_k|\alpha\rangle\} \tag{5}$$

$$+ i\sum_{i,k=1}^r \langle [\hat{\xi}_k, \hat{\xi}_i] \rangle \sum_{\alpha<\beta} (p_\alpha + p_\beta)(h_\beta - h_\alpha)\Im\{\langle\alpha|\hat{X}_k|\beta\rangle\langle\beta|\hat{X}_i|\alpha\rangle\}, \tag{6}$$

where $\{...,...\}_+$ denotes anti-commutator, $\Re$ and $\Im$ stand for the real and imaginary parts, respectively. $\{|\alpha\rangle\}$ is the common eigenbase of $\rho_S(0)$ and $\bar{H}_S(0)$, and p_α, h_α are the corresponding eigenvalues: $p_\alpha = e^{-\beta h_\alpha}/\sum_\alpha e^{-\beta h_\alpha}$. For $r = 1$, where only the

term (5) survives, one can show that $W > 0$ because $\hat{\rho}_S(0)$ is a decreasing function of $\hat{H}_S$. Since the thermodynamic contribution to the work is negligible small, the amount of dissipated work predicted by the second law, is *enhanced* in the mesoscopic domain. The contribution (6) reflects an effect of interference between different types of work (channels), since it contains non-diagonal terms over i, k and becomes zero with $\langle [\hat{\xi}_k, \hat{\xi}_i] \rangle$. Now assume that the temperature of the system S is so high (fluctuations are strong) that $p_\alpha \simeq p_\beta$ for all the β and α that are connected by matrix elements of $\hat{X}_k$ and $\hat{X}_i$. Then the contribution of (5) can be disregarded, and provided that $\langle [\hat{\xi}_k, \hat{\xi}_i] \rangle \neq 0$ and $\Im\{\langle \alpha | \hat{X}_k | \beta \rangle \langle \beta | \hat{X}_i | \alpha \rangle\} \neq 0$, the sign of (6) can be made negative. Thus the many-source situation can lead to a violation of the second law. The above results also apply to the classical limit, where Poisson brackets take the place of commutators.

Let us then examine the thermalization process which takes place for $t > \tau$. It is straightforward to show that due to the taken definition of cyclic processes, marginal density operator of S is the equilibrium distribution (2), both at the preparation time $t = 0$ and after the interaction with the thermal bath that takes place for $t > \tau$.

Exactly solvable models. The result above on the violation and enhancement of the Thomson's formulation of the second law do not use specific properties of the involved systems S and W, but unfortunately are essentially restricted to short times. Moderate and long times can be explored by help of exactly solvable models for S and W. To this end we investigated two such models [8, 9]. The first of them is Jaynes-Cummings model describing interaction of two-level atom or spin and a single mode of radiation [10]. As for our purposes, we can naturally identify the two-level atom as the system S, and the electromagnetic field - which is controlled externally and admits a well-defined macroscopic limit - as the source W. The Hamiltonian of the model reads

$$\hat{H}_S = \frac{1}{2}\hbar\omega\hat{\sigma}_z, \quad \hat{H}_W = \hbar\omega\hat{a}^\dagger\hat{a}, \quad \hat{H}_I = \hbar g(\hat{\sigma}_+\hat{a} + \hat{\sigma}_-\hat{a}^\dagger), \tag{7}$$

where $\hat{a}$ ($\hat{a}^\dagger$) is the annihilation (creation) operators of the field, $\hat{\sigma}_x$, $\hat{\sigma}_y$, $\hat{\sigma}_z$, $\hat{\sigma}_{+,-} = (\hat{\sigma}_x \pm i\hat{\sigma}_y)/2$ are the Pauli matrices of the spin, and $g > 0$ is the coupling constant. The initial state of the spin is thermal with temperature T and that of the mode is described by a coherent state $|\alpha\rangle$ with an amplitude α. With help of the known exact solution of this model we found some new mechanisms of the Thomson's law violation in the mesoscopic regime; for a detailed discussion see [8].

In the second exactly solvable model [9] the systems S and W are two ensembles of N independent classical harmonic oscillators, coupled one-to-one via their coordinates x and y. Like in an ideal gas, one may describe only one set of them, with Hamiltonians

$$H_S = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega^2 x^2, \quad H_W = \frac{1}{2}M\dot{y}^2 + \frac{1}{2}M\Omega^2 y^2, \quad H_I = -gxy, \tag{8}$$

The x-oscillator, associated with S, starts its evolution from a thermal state at a temperature T. The y-oscillator starts from a macroscopic state with a well-defined initial coordinate $y(0)$ and velocity $\dot{y}(0)$. It plays the role of the work-source W. This model not only allows to discuss limits of the second law for one closed cycle, but actually admits rather detailed many-cycle considerations [9]. We determined conditions under which

an infinity of cycles with constant yield is possible. This is a mesoscopic 'perpetuum mobile' that may cool the macroscopic bath B [9].

Clausius principle and its microscopic foundations.

Finally, we shortly analyze the Clausius principle: heat goes from higher temperature to lower one. To put this statement on the microscopic ground, one considers two interacting systems 1 and 2 with the total Hamiltonian [1, 2, 11, 12]: $\hat{H}_{\text{tot}} = \hat{H}_1 + \hat{H}_2 + \hat{H}_I$, where $\hat{H}_1$ ($\hat{H}_2$) is the Hamiltonian of the subsystem 1 (2), and $\hat{H}_I$ is the interaction Hamiltonian. The initial state consists of two equilibrium subsystems at different temperatures, $\rho_{\text{tot}}(0) = \rho_1(0) \otimes \rho_2(0)$, where $\rho_k(0) = \exp[-H_k/T_k]/Z_k$ for $k = 1, 2$.

Let us denote $\Delta U_1 = \langle \hat{H}_1 \rangle_t - \langle \hat{H}_1 \rangle_0$, $\Delta U_2 = \langle \hat{H}_2 \rangle_t - \langle \hat{H}_2 \rangle_0$ and $\Delta U_I = U_I(t) - U_I(0^+)$. It was rigorously proven that [11]:

$$\frac{\Delta U_1}{T_1} + \frac{\Delta U_2}{T_2} \geq 0, \tag{9}$$

whereas correct but less rigorous arguments towards its validity are provided in [1]. After the coupling between the two subsystems has been turned on, the total energy is conserved. Thus we get from (9): $(T_2 - T_1)\, \Delta U_1 \geq T_1 \Delta U_I$. In case $T_2 > T_1$ and $\Delta U_I \geq 0$ this inequality proves the Clausius principle. If $\hat{H}_I$ commutes with $\hat{H}_1$ and $\hat{H}_2$, U_I is conserved, implying that $(T_2 - T_1)\, \Delta U_1 \geq 0$, as proven by Reents [12].

When $T_2 > T_1$ but $\Delta U_I < 0$ there may, in principle, be a flow from low to high temperature, but it is bounded if ΔU_I is bounded. In particular, such a current cannot be constant and everlasting.

REFERENCES

1. L.D. Landau and E.M. Lifshitz, *Statistical Physics, I*, (Pergamon Press Oxford, 1978).
2. R. Balian, *From Microphysics to Macrophysics*, volumes I and II, (Springer, 1992).
3. F. Schlögl, Phys. Rep. **62**, 268 (1980).
4. A. Lenard, J. Stat. Phys **19**, 575, (1978). W. Pusz, L. Woronowicz, Comm. Math. Phys., **58**, 273, (1978). G.N. Bochkov and Yu.E. Kuzovlev, Physica A, **106**, (1981) 443. J. Kurchan, cond-mat/0007360.
5. A.E. Allahverdyan and Th.M. Nieuwenhuizen, Physica A **305**, 542 (2002).
6. A.E. Allahverdyan and Th.M. Nieuwenhuizen, Phys. Rev. Lett. **85** (2000) 1799; Phys. Rev. E **64**, 056117 (2001); Th.M. Nieuwenhuizen and A.E. Allahverdyan, Phys. Rev. E, to appear (2002) 52 pages, cond-mat/0011389; Th.M. Nieuwenhuizen and A.E. Allahverdyan, *Quantum Brownian motion and its conflict with the second law*, these proceedings.
7. C. Pombo, A.E. Allahverdyan and Th.M. Nieuwenhuizen, these proceedings; A.E. Allahverdyan and Th.M. Nieuwenhuizen, cond-mat/0201408.
8. A.E. Allahverdyan, R. Balian and Th.M. Nieuwenhuizen, submitted to Phys. Rev. E (2002).
9. Th.M. Nieuwenhuizen, A.E. Allahverdyan and R. Balian, preprint, (July, 2002).
10. M. Scully and S. Zubairy, *Quantum Optics*, (Cambridge University Press, 1997).
11. H. Tasaki, cond-mat/0009244.
12. G. Reents, these proceedings.